

# CONTENTS

<b>5 Data Reduction</b>	<b>2</b>
5.1 Sufficient Statistics . . . . .	2
5.2 Factorization Theorem . . . . .	5
5.3 Rao-Blackwell . . . . .	10
5.4 Completeness . . . . .	13
5.5 Exponential Class . . . . .	17
5.6 Lehmann-Scheffe . . . . .	20
5.7 Basu Theorem . . . . .	25

## 5 Data Reduction

### 5.1 Sufficient Statistics

#### Definition 1. Jointly Sufficient Statistics

Let  $\mathbf{X} = (X_1, \dots, X_n)$  have joint pdf  $f(\mathbf{x}, \boldsymbol{\theta})$ , and let  $S = (S_1, \dots, S_k)$  be a  $k$ -dimensional statistic. Then  $S_1, \dots, S_k$  is a set of **jointly sufficient statistics** for  $\boldsymbol{\theta}$  if for any other vector of statistics,  $\mathbf{T}$ , the conditional pdf of  $\mathbf{T}$  given  $\mathbf{S} = \mathbf{s}$ , denoted by  $f_{\mathbf{T}|\mathbf{S}}(t)$ , does not depend on  $\boldsymbol{\theta}$ . In the one-dimensional case, we simply say that  $S$  is a **sufficient statistic** for  $\theta$ .

**Example 1.**

Let  $X_1, \dots, X_n$  be a random sample from a Bernoulli distribution. Show that  $S = \sum X_i$  is a sufficient statistic for  $\theta$  by definition.

**Example 2.**

Consider a random sample from an exponential distribution,  $X_i \sim EXP(\theta)$ . Show that  $S = \sum X_i$  is a sufficient statistic for  $\theta$  by definition.

## 5.2 Factorization Theorem

### Theorem 1. Factorization Criterion

If  $X_1, \dots, X_n$  have joint pdf  $f(x_1, \dots, x_n; \theta)$ , and if  $S = (S_1, \dots, S_k)$ , then  $S_1, \dots, S_k$  are jointly sufficient for  $\theta$  if and only if

$$f(x_1, \dots, x_k; \boldsymbol{\theta}) = g(\mathbf{s}; \boldsymbol{\theta})h(x_1, \dots, x_n)$$

where  $g(\mathbf{s}; \boldsymbol{\theta})$  does not depend on  $x_1, \dots, x_n$ , except through  $s$ , and  $h(x_1, \dots, x_n)$  does not involve  $\theta$ .

### Example 3.

Consider a random sample from a Gamma distribution,  $X_i \sim \text{Gamma}(\alpha = 3, \theta)$ . Show that  $S = \sum X_i$  is a sufficient statistic for  $\theta$  by factorization theorem.

**Definition 2.** If  $A$  is a set, then the indicator function of  $A$ , denoted by  $I_A$  is defined as

$$I_A = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$

**Example 4.** Consider a random sample from a uniform distribution,  $X_i \sim U(0, \theta)$ , where  $\theta$  is unknown. Find a sufficient statistic for  $\theta$ .

**Example 5.** Consider a random sample from a uniform distribution,  $X_i \sim U(\theta, \theta + 1)$ . Notice that the length of the interval is one unit, but the endpoints are assumed to be unknown. Show that  $S_1 = X_{1:n}$  and  $S_2 = X_{n:n}$  are jointly sufficient for  $\theta$ .

**Example 6.**

Consider a random sample from a normal distribution,  $X_i \sim N(\mu, \sigma^2)$ , where both  $\mu$  and  $\sigma^2$  are unknown. Show that  $S_1 = \sum X_i$  and  $S_2 = \sum X_i^2$  are jointly sufficient for  $\boldsymbol{\theta} = (\mu, \sigma^2)$ .

**5.3 Rao-Blackwell**

**Theorem 2. Rao-Blackwell** Let  $X_1, \dots, X_n$ , have joint pdf  $f(x_1, \dots, x_n; \boldsymbol{\theta})$ , and let  $S = (S_1, \dots, S_k)$  be a vector of jointly sufficient statistics for  $\boldsymbol{\theta}$ . If  $T$  is any unbiased estimator of  $\tau(\boldsymbol{\theta})$ , and if  $T^* = E(T|S)$ , then

1.  $T^*$  is an unbiased estimator of  $\tau(\boldsymbol{\theta})$
2.  $T^*$  is a function of  $S$ , and
3.  $V(T^*) \leq V(T)$  for every  $\boldsymbol{\theta}$ , and  $V(T^*) < V(T)$  for some  $\boldsymbol{\theta}$  unless  $T^* = T$  with probability 1.

**Example 7.**

Let  $X_1, X_2, \dots, X_n$  be random sample of size  $n$  from a Poisson distribution with unknown mean  $\lambda$ . Find the UMVUE of  $\theta = \frac{\lambda^x e^{-\lambda}}{x!}$ .

**Example 8.**

Let  $X_1, X_2, \dots, X_n$  be random sample of size  $n$  from an Exponential distribution with unknown mean  $\theta$ . Find the UMVUE of  $\gamma = 1 - e^{-t/\theta}$  using Rao-Blackwell theorem.

## 5.4 Completeness

**Definition 3. Completeness** A family of density functions  $\{f_{\mathbf{T}}(t; \boldsymbol{\theta}); \boldsymbol{\theta} \in \Omega\}$  is called complete if  $E[u(T)] = 0$  for all  $\boldsymbol{\theta} \in \Omega$  implies  $u(T) = 0$  with probability 1 for all  $\boldsymbol{\theta} \in \Omega$ .

A sufficient statistic the density of which is a member of a complete family of density functions will be referred to as a **complete sufficient statistic**.

## Example 9.

Let  $X_1, \dots, X_n$  denote a random sample from a Geometric distribution,  $X_i \sim \text{Geometric}(p)$ . Show that  $S = \sum X_i$  is the complete sufficient statistic for  $p$ .

**Example 10.**

Let  $X_1, \dots, X_n$  be iid  $N(\theta, a\theta^2)$ , where  $a$  is known constant and  $\theta > 0$ . Show that the family of distribution is not complete.

**Example 11.**

Consider a random sample of size  $n$  from a uniform distribution  $X \sim U(0, \theta)$ . Show that  $S = X_{n:n}$  is a complete sufficient statistic for  $\theta$ .



## 5.5 Exponential Class

**Definition 4. Exponential Class** A density function is said to be a member of the regular exponential class if it can be expressed in the form

$$f(\mathbf{x}; \boldsymbol{\theta}) = c(\boldsymbol{\theta})h(\mathbf{x})e^{-\sum_{j=1}^n q_j(\boldsymbol{\theta})t_j(\mathbf{x})}, \mathbf{x} \in A$$

and zero otherwise, where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  is a vector of  $k$  unknown parameters, if the parameter space has the form

$$\Omega = \{\boldsymbol{\theta} | a_i \leq \theta_i \leq b_i, i = 1, \dots, k\}$$

(note that  $a_i = -\infty$  and  $b_i = \infty$  are permissible values), and if it satisfies regularity conditions 1, 2, and 3(a) or 3(b) given by:

1. The set  $A = \{x : f(x; \boldsymbol{\theta}) > 0\}$  does not depend on  $\boldsymbol{\theta}$ .
2. The functions  $q_j(\boldsymbol{\theta})$  are nontrivial, functionally independent, continuous functions of the  $\boldsymbol{\theta}$ .
- 3.(a) For a continuous random variable, the derivatives  $t_j(x)$  are linearly independent contin-

uous functions of  $x$  over  $A$ .

- (b) For a discrete random variable, the  $t_j(x)$  are nontrivial functions of  $x$  on  $A$ , and none is a linear function of the others.

For convenience, we will write that  $f(x; \boldsymbol{\theta})$  is a member of  $REC(q_1, \dots, q_k)$  or simply  $REC$ .

**Theorem 3.**  $X_1, \dots, X_n$ , is a random sample from a member of the regular exponential class  $REC(q_1, \dots, q_k)$ , then the statistics

$$S_1 = \sum_{i=1}^n t_1(X_i), \dots, S_k = \sum_{i=1}^n t_k(X_i)$$

are a minimal set of complete sufficient statistics for  $\theta_1, \dots, \theta_k$ .

**Example 12.**

Show that  $X \sim \text{Gamma}(\alpha, \theta)$  belong to the regular exponential class, and use this information to find complete sufficient statistics based on a random sample  $X_1, \dots, X_n$ .

**5.6 Lehmann-Scheffe**

**Theorem 4. Lehmann-Scheffe** Let  $X_1, \dots, X_n$  have joint pdf  $f(x_1, \dots, x_n; \boldsymbol{\theta})$ , and let  $S$  be a vector of jointly complete sufficient statistics for  $\boldsymbol{\theta}$ . If  $T^* = t^*(S)$  is a statistic that is unbiased for  $\tau(\boldsymbol{\theta})$  and a function of  $S$ , then  $T^*$  is a UMVUE of  $\tau(\boldsymbol{\theta})$ .

**Example 13.**

Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a random sample from a Weibull distribution,  $X \sim WEI(\tau = 4, \theta)$ . Find the UMVUE for  $\theta$ .

**Example 14.**

Let  $X_1, \dots, X_{30}$  be a random sample from a distribution with probability density function (p.d.f.)

$$f(x) = \frac{\theta^4}{\Gamma(4)} x^3 e^{-\theta x} I(0, \infty), \theta > 0.$$

- (a) Show that the p.d.f. of  $X$  belongs to the regular exponential family.
- (b) Find a complete and sufficient statistic for  $\theta$ .
- (c) Find the UMVUE for  $V(X_1)$ .
- (d) Find the UMVUE for  $\theta$ .

**Example 15.**

Let  $X_1, X_2, \dots, X_n$  be random sample of size  $n$  from  $f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, x = 0, 1, \dots, \infty$ . Find the uniformly minimum variance unbiased estimator (UMVUE) of  $g(\theta) = \frac{e^{-\theta}\theta^4}{4!}$ .

**Example 16.**

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution with pdf

$$f(x; \theta) = 7\theta x^{7\theta-1}.$$

Find the UMVUE of  $\theta$ ,

## 5.7 Basu Theorem

**Theorem 5. Basu** Let  $X_1, \dots, X_n$  have joint pdf  $f(x_1, \dots, x_n; \theta)$ ;  $\theta \in \Omega$ . Suppose that  $S = (S_1, \dots, S_k)$  where  $S_1, \dots, S_k$  are jointly complete sufficient statistics for  $\theta$ , and suppose that  $T$  is any other statistic. If the distribution of  $T$  does not involve  $\theta$ , then  $S$  and  $T$  are stochastically independent. In this case,  $T$  is called an ancillary statistic.

**Example 17.** Consider a random sample of size  $n$  from a normal distribution  $X \sim N(\mu, \sigma^2)$ . Use Basu Theorem to show that  $\bar{X}$  and  $S^2$  are independent.

**Example 18.**

Consider a random sample of size  $n$  from a two-parameter exponential distribution,  $X_i \sim EXP(1, \eta)$ .

- (a) Show that  $T(X) = X_{1:n}$  is complete and sufficient for  $\eta$ .
- (b) Find the UMVUE of  $\eta$ .
- (c) Find the UMVUE of  $\eta^2$ .
- (d) Use Basu's Theorem to show that  $X_{1:n}$  and  $S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$  are independent.